JOURNAL OF COMPLEXITY **16**, 390–410 (2000)
doi:10.1006/jcom.2000.0541, available online at http://www.idealibrary.com on IDEAL®

Average Approximations and Moments of Measures

Vladimir Bogachev[1]

Department of Mechanics and Mathematics, Moscow State University, 119899 Moscow, Russia

Received October 1, 1996

We investigate average approximations of infinite dimensional mappings and related problems connected with moments of measures on linear spaces. A conjecture stated by J. F. Traub and A. G. Werschulz (1994, *Math. Intelligencer* **16**, 42–48) is settled. Several positive results concerning average approximations of Banach space valued mappings are obtained. Some related open problems are discussed. © 2000 Academic Press

INTRODUCTION

In recent years there has been a growing interest in average solutions for ill-posed problems [TWW88, Wer87, Wer91] and, more generally, in the estimation of approximating algorithms by means of measures on suitable (typically infinite dimensional) spaces. A detailed discussion of this direction can be found in the references cited, so we shall describe only the concrete situation we deal with below. Let X be a separable Banach space with a Borel probability measure μ and let $F: X \to X$ be a μ-measurable mapping. Suppose we are given a certain class Λ of real functions on X interpreted as permissible information functions. The element $N(x) = (\lambda_1(x), ..., \lambda_n(x))$, where $\lambda_i \in \Lambda$, is called an input or an information, and the mapping N is called an information operator. Here the number n may depend on x and the functions λ_i may be constructed adaptively; i.e., recursive inputs $N(x)$ of the following form are allowed,

$$\{\lambda_1(x), \lambda_2(x, \lambda_1(x)), ..., \lambda_{n(x)}(x, \lambda_1(x), ..., \lambda_{n(x)-1}(...))\},$$

[1] Partially supported by the Russian Foundation of Fundamental Research Grants 96-15-96865 and 97-01-00932, Grant DFG 436 RUS 113/343/0(R), and INTAS Projects 94-0378 and 95-0099.

0885-064X/00 $35.00

where for any k, $\lambda_k(x, y_1, ..., y_{k-1})$ is a function on $X \times \mathbb{R}^{k-1}$ which is measurable in x and Borel in the other arguments. In addition, we assume that substitution of λ_i for y_i yields a measurable function. The operator N is called *nonadaptive* if only inputs with λ_k independent of $(y_1, ..., y_{k-1})$ are allowed and n is fixed. An algorithm is a mapping φ that associates an element of X to every input $N(x)$. We refer to [TWW88] concerning the terminology.

In the average setting, the quality of an algorithm φ is measured by

$$\varepsilon(\varphi, N) = \int_X \|F(x) - \varphi(N(x))\|^2 \, \mu(dx). \tag{0.1}$$

For some problems also $\|F - \varphi \circ N\|_{L^p(\mu, X)}$ can be useful for estimating the quality of algorithms.

In this setting the following natural questions arise:

(i) Are there an information operator N and an algorithm φ such that $\varepsilon(\varphi, N) < \infty$?

(ii) If "yes," then what is the infimum of such quantities over all admissible φ and N or over all φ for a fixed N?

If the answer to (i) is "yes" and $\varepsilon(\varphi, N)$ in (0.1) can be made arbitrarily small, then the corresponding problem is called in [TW94] *solvable on the average*. If the answer to (i) is "no," then the problem is called *unsolvable on the average*. Finally, if the answer to (i) is "yes," but $\varepsilon(\varphi, N) < \infty$ cannot be made smaller than some $\varepsilon_0 > 0$, then the problem is called *weakly solvable on the average*.

Note that if F is invertible μ-a.e., then replacing the initial measure μ by its image $\nu = \mu \circ F^{-1}$ defined by $\nu(B) = \mu(F^{-1}(B))$ and the functions λ_i by $\zeta_i = \lambda \circ F^{-1}$, we reduce the problem to the case where F is the identity map, since

$$\varepsilon(\varphi, N) = \int_X \|F(x) - \varphi(N(x))\|^2 \, \mu(dx) = \int_X \|z - \varphi(N \circ F^{-1}(z))\|^2 \nu \, (dz).$$

This observation links the problem above with the problem of existence of finite moments of measures.

Note that in applications it is typical to have for Λ a certain linear space of μ-measurable linear functions on X (see Definition 1.1 below); e.g., Λ can coincide with X^*. Below we consider only this case.

It has been shown by Traub and Werschulz [TW94] that if F is a μ-measurable linear operator, it can happen that $\int \|F(x)\|^2 \, \mu(dx) = \infty$, but the problem is solvable on the average for the class of nonadaptive informations; in particular, there exist an algorithm φ and a nonadaptive information operator N such that the quantity in (0.1) is finite. Note that if X is a Hilbert space and μ has finite second moment, i.e., $\int \|x\|^2 \mu(dx) < \infty$, then the problem is solvable on the average for $F = I$ in the class of nonadaptive informations. If μ is a Gaussian measure, the same is true for arbitrary X and any μ-measurable linear operator F. In this connection, it has been conjectured by Traub and Werschulz [TW94] that there exist measures μ such that for the operator $F = I$ the problem is unsolvable on the average. We prove that this conjecture is true. Moreover, in our example μ is the mixture of homothetic images of a centered Gaussian measure. A positive result proved below states that if μ on X has finite moment of order p, then there is a separable reflexive Banach space E compactly embedded into X such that $\mu(E) = 1$ and $\int \|x\|_E^p \mu(dx) < \infty$. This result extends a theorem from [Ost80], where E was the dual to a separable Banach space. Below we discuss possible applications of this result to the study of average approximations. Some results from [M90] are improved. Related problems have been discussed by many authors; see, e.g., [H90, M90, TWW88, TW94, Was86, WW84, Wer87, Wer91], and the references therein.

Helpful comments and suggestions of two very thorough referees are gratefully acknowledged.

1. NOTATION AND TERMINOLOGY

Let X be a separable Banach space with the topological dual X^*. Below, *a measure μ on X* means a finite nonnegative countably additive measure μ defined on the σ-field $\mathscr{B}(X)$ of all Borel subsets of X. For such a measure, we denote by $\mathscr{B}_\mu$ the Lebesgue completion of $\mathscr{B}(X)$ with respect to μ. A mapping F from X to a topological space Y is called μ-measurable if $F^{-1}(B) \in \mathscr{B}_\mu$ for all $B \in \mathscr{B}(Y)$.

We say that μ has finite moment of order $p > 0$ if $\int \|x\|^p \mu(dx) < \infty$.

We say that $E \subset X$ is a Banach space compactly embedded into X if E is a linear subspace of X endowed with a stronger norm $\|\cdot\|_E$ such that $(E, \|\cdot\|_E)$ is a Banach space, and its unit ball is relatively compact in X. For example, $C^1[0, 1] \subset C[0, 1]$ is a compactly embedded Banach space.

DEFINITION 1.1. A μ-measurable map F on X with values in a separable Banach space Y is said to be a μ-measurable linear operator if it admits an equivalent modification that is linear on X in the usual sense (such a modification is said to be a proper measurable linear operator). Measurable linear operators with values in $\mathbb{R}^1$ are called measurable linear functionals.

Note that if there is a sequence $\{f_n\}$ of continuous linear functionals on X convergent to f in measure, then f is a μ-measurable linear functional. Indeed, then a subsequence of $\{f_n\}$ (denoted by the same indices) converges to f almost everywhere. The domain of its convergence L is clearly a measurable linear subspace of full measure. Now we redefine f, which is given by $\lim_{n\to\infty} f_n$ on L, on the complement of L using a Hamel basis in X; thus we get a proper linear version. The converse is not true: in general, there exist μ-measurable linear functionals f that are not limits of sequences of continuous linear functionals convergent in measure (see, e.g., [K78, Sm83]). However, both definitions are equivalent for Gaussian measures (see [B98, Sect. 2.10; Bor76]). A typical example of a measurable linear functional that is not continuous is a stochastic integral with a nonrandom integrand with respect to the Brownian motion. We shall find other examples below. The reason why not necessarily proper linear versions are admissible is that it is not always possible to choose simultaneously proper linear versions of an uncountable family of measurable linear functionals.

Recall that a probability measure μ on X is called Gaussian if for every $f \in X^*$ the induced measure $\mu \circ f^{-1}$ on $\mathbb{R}^1$ is Gaussian, i.e., is either a Dirac measure at some point a or admits a Gaussian density $(2\pi\sigma)^{-1/2} \exp(-(t-a)^2/(2\sigma))$. If all these induced measures are centered (i.e., $a=0$), then μ is called centered.

If μ is a centered Gaussian measure, then the collection Λ_μ of all μ-measurable linear functionals coincides with the closure of X^* in $L^2(\mu)$. Letting

$$|h|_\mu = \sup\left\{ f(h) : f \in X^*, \int f(x)^2\, \mu(dx) \leqslant 1 \right\},$$

one defines the Cameron–Martin space of μ (also called its reproducing kernel Hilbert space) by

$$H(\mu) = \{ h \in X : |h|_\mu < \infty \}.$$

The Cameron–Martin space becomes a Hilbert space with the norm $|\cdot|_\mu$.

The most typical example of a centered Gaussian measure with the infinite dimensional support is the countable product of the standard Gaussian measures on $\mathbb{R}^1$; in fact, this example is unique up to a measurable linear isomorphism. This product μ is originally defined on $\mathbb{R}^\infty$ (the countable product of real lines), but can be restricted to any Banach or Hilbert subspace X of $\mathbb{R}^\infty$ with full measure. For example, we can take

$$X = \left\{ (x_n) : \|x\|^2 = \sum_{n=1}^{\infty} n^{-2} x_n^2 < \infty \right\}, \qquad (x, y)_X = \sum_{n=1}^{\infty} n^{-2} x_n y_n. \quad (1.2)$$

Note that in this example $H(\mu) = l^2$ and all μ-measurable linear functionals are described by the formula

$$\xi = \sum_{n=1}^{\infty} c_n x_n, \qquad (c_n) \in l^2,$$

where the series converges in $L^2(\mu)$.

It is worth noting that if a linear function f on X is measurable with respect to every Gaussian measure on X, then f is continuous (see [Bor76; B98, Proposition 3.11.12]). In particular, a discontinuous linear functional cannot be measurable with respect to all Borel measures; hence the collection of μ-measurable linear functions depends on μ in an essential way.

A probability measure ν on X is said to be pre-Gaussian if there is a Gaussian measure μ on X such that

$$\int_X f(x)\, \nu(dx) = \int_X f(x)\, \mu(dx),$$

$$\int_X f(x)\, g(x)\, \nu(dx) = \int_X f(x)\, g(x)\, \mu(dx), \qquad \forall f, g \in X^*.$$

2. THE TRAUB–WERSCHULZ PROBLEM

If X is a Hilbert space (or a Banach space with a Schauder basis; see [Sin81]) and μ is a measure on X with finite second moment, then taking a basis $\{e_n\}$ in X and considering the corresponding projections P_n to the linear spans of $e_1, ..., e_n$, we get immediately

$$\int_X \|x - P_n x\|^2\, \mu(dx) \to 0 \qquad \text{as} \qquad n \to \infty.$$

Indeed, $\|x - P_n x\| \to 0$ pointwise and $\sup_n \|P_n\| < \infty$; hence Lebesgue's dominated convergence theorem applies. Therefore, for any μ-measurable linear operator $F\colon X \to X$ with $\int_X \|Fx\|^2\, \mu(dx) < \infty$, applying the previous relationship to the measure $\nu = \mu \circ F^{-1}$ (the image of μ under F) and letting $Q_n x = P_n F x$, one has

$$\int_X \|Fx - Q_n x\|^2\, \mu(dx) \to 0 \qquad \text{as} \qquad n \to \infty.$$

Thus, we get a known result (see, e.g., [TW94]) stating that well-posed on the average problems are solvable on the average. It is natural to ask whether the same is true for general Banach spaces not necessarily having Schauder bases. The next result shows that composing finite dimensional linear operator with nonlinear mappings, one gets a similar result for arbitrary Banach spaces. In the next section we shall discuss linear approximations.

PROPOSITION 2.1. *Let μ be a measure on a separable Banach space X and let $F: X \to X$ be a μ-measurable linear operator such that $\int \|Fx\|^2 \, \mu(dx) < \infty$. Then, for every $\varepsilon > 0$, there exist μ-measurable linear functionals l_i, $i = 1, ..., n$, and a continuous mapping φ: $\mathbb{R}^n \to X$ such that*

$$\int_X \|Fx - \varphi(l_1(x), ..., l_n(x))\|^2 \, \mu(dx) < \varepsilon.$$

Proof. As explained above, it suffices to prove our claim for the identity operator and measure $v = \mu \circ F^{-1}$ and then replace l_i found for v by $l_i \circ F$. By a classical result, X is linearly isometric to a closed linear subspace of $C[0, 1]$. According to [V94, Theorem 3 (part (c) of which applies since $C[0, 1]$ is separable)], there is an equivalent norm on $C[0, 1]$ such that X becomes a Chebyshev set for this norm and has a continuous metric projection. This means that if one lets Z be $C[0, 1]$ with this new norm, for every $z \in Z$ there is a unique $\psi(z) \in X$ with the property $\|\psi(z) - z\|_Z = \inf_{x \in X} \|x - z\|_Z$, and, in addition, $\psi: Z \to X$ is continuous. Since $C[0, 1]$ with the usual norm has a Schauder basis $\{e_n\}$, by the equivalence of the two norms, $\{e_n\}$ is a Schauder basis in Z. Hence, letting P_n be the corresponding projections, we have $P_n z \to z$ in Z for every $z \in Z$, and $\sup_n \|P_n\|_{L(Z)} \leqslant C < \infty$. Clearly,

$$\|x - \psi(P_n x)\|_Z \leqslant \|x - P_n x\|_Z, \qquad \forall x \in X.$$

By the equivalence of the initial norm of X to that induced from Z, we have $\|y\| \leqslant k \, \|y\|_Z$ for all $y \in X$. Therefore, by Lebesgue's dominated convergence theorem

$$\int_X \|x - \psi(P_n x)\|^2 \, \mu(dx) \leqslant k^2 \int_X \|x - \psi(P_n x)\|_Z^2 \, \mu(dx)$$

$$\leqslant k^2 \int_X \|x - P_n x\|_Z^2 \, \mu(dx) \to 0,$$

since

$$\|x - P_n x\|_Z \leqslant \|x\|_Z + \|P_n x\|_Z \leqslant (1+C)\,\|x\|_Z \leqslant k(1+C)\,\|x\| \in L^2(\mu).$$

Since there is a linear isomorphism $T_n\colon P_n(X) \to \mathbb{R}^n$, we get a continuous operator $T_n \circ P_n\colon X \to \mathbb{R}^n$, which can be written as n linear functionals $l_i \in X^*$. The mapping $\varphi := \psi \circ T_n^{-1}\colon \mathbb{R}^n \to X$ has the required properties. ∎

Note that the arguments based on metric projections are standard in the approximation theory; approximate metric projections were used in [M90] for constructing uniform approximations. Note also that if one admits non-linear functions l_i, then, according to [M90, Theorem 12], it suffices to take $n=1$. Easy examples show that this is not always possible for linear l_i.

As shown by Traub and Werschulz [TW94], it may happen that μ has no finite second moment, but it is still possible to approximate the identity mapping by a mapping depending on finitely many coordinates in the following sense: there exist μ-measurable linear functionals $l_1, ..., l_n$, and a Borel mapping $\varphi\colon \mathbb{R}^n \to X$ (in [TW94] the mapping is even continuous) such that

$$\int_X \|x - \varphi(l_1(x), ..., l_n(x))\|^2\,\mu(dx) < \infty.$$

It has been conjectured by Traub and Werschulz [TW94] that there exist measures without such approximations. We shall construct an example confirming Traub–Werschulz's conjecture and discuss several related questions.

Example 2.2. Let X be the Hilbert space defined in (1.2) and let γ be the countable product of the standard Gaussian measures on $\mathbb{R}^1$. Then γ is a centered Gaussian measure on X. Define a probability measure μ on X by

$$\mu(A) = \int_0^\infty \gamma(A/t)\,p(t)\,dt,$$

where p is any positive probability density on $(0, \infty)$ such that $\int_0^\infty t^2 p(t)\,dt = \infty$. Then, for any finite collection of μ-measurable linear functionals $l_1, ..., l_n$, and any Borel mapping $\varphi\colon \mathbb{R}^n \to X$, one has

$$\int_X \|x - \varphi(l_1(x), ..., l_n(x))\|^2\,\mu(dx) = \infty.$$

Proof. Below we shall make use of several standard facts in the theory of Gaussian measures (see, e.g., [B98, Bor76]). Let us put $\gamma_t(A) = \gamma(A/t)$. For every positive t, the measure γ_t is a centered Gaussian measure on X. Its Cameron–Martin space H_t coincides as a set with l^2, but the inner product is given by $(a, b)_{H_t} = t^{-2} \sum_{n=1}^{\infty} a_n b_n$.

First of all, let us prove that any measurable linear functional l on (X, μ) can be written as

$$l(x) = \sum_{n=1}^{\infty} c_n x_n, \tag{2.3}$$

where $(c_n) \in l^2$ and the series converges μ-almost everywhere. That is, the description of measurable linear functionals is the same as in the case where μ is Gaussian. Indeed, by definition l is μ-measurable and linear on a linear subspace L of full μ-measure. It is easy to prove that for a.e. t one has, the set L and the functional l are γ_t-measurable and $\gamma_t(L) = 1$; in fact, all the measures γ_t are mutually singular, but nevertheless have equal collections of measurable linear functionals and measurable linear subspaces. Indeed, if $A \in \mathscr{B}_\mu$, then one can find two Borel sets B_1 and B_2 such that $B_1 \subset A \subset B_2$ and $\mu(B_2 \backslash B_1) = 0$. Then, $\gamma_t(B_2 \backslash B_1) = 0$ for almost all t, whence $A \in \mathscr{B}_{\gamma_t}$ for such t.

Since, by definition, l has a proper linear version, we shall work with this version. Let $\tau > 0$ be such that l is γ_τ-measurable. By virtue of a classical result (see [B98, Corollary 2.10.10; Bor76]), there is a sequence (c_n) such that $\sum_{n=1}^{\infty} c_n^2 < \infty$ and the series in (2.3) converges γ_τ-a.e. Moreover, since l is linear in the usual sense, one has $c_n = l(e_n)$, where $\{e_n\}$ is the standard orthonormal basis in l^2. Clearly, then the series in (2.3) converges γ_t-a.e. for all $t > 0$. The domain L of its convergence is automatically a Borel linear subspace in X, and it follows from our reasoning that it has full μ-measure, since it has full γ_t-measure for all $t > 0$. The function ξ defined on L by $\xi(x) = \sum_{n=1}^{\infty} c_n x_n$ is a μ-measurable functional which is linear on L. Since γ_t-measurable proper linear functionals are uniquely determined by their values on $\{e_n\}$ (see [B98, Theorem 2.10.7; Bor76]) and $l(e_n) = \xi(e_n)$, we have $l = \xi$ γ_t-a.e. for every t such that l is γ_t-measurable. Therefore, $l = \xi$ μ-a.e. We shall deal with the versions of μ-measurable linear functionals l given by the sums of the series in (2.3) on the corresponding domains of convergence. With this convention, such versions are well defined and measurable simultaneously with respect to all measures γ_t, $t > 0$, although all these measures are mutually singular and also singular with respect to μ.

Now let $\varphi: \mathbb{R}^n \to X$ be a Borel mapping and let $l_1, \ldots, l_n$ be μ-measurable linear functionals on X. The space Λ of all μ-measurable linear functionals on X is equipped with the Hilbertian inner product $(l, k)_2 = \sum_{n=1}^{\infty} c_n b_n$, where l and k are given in representation (2.3) by the sequences (c_n) and (b_n), respectively. Applying the orthogonalization and normalization to the elements $l_1, ..., l_n$, we get some μ-measurable linear functionals $\xi_1, ..., \xi_n$ (maybe, in a number less than n) which are mutually orthogonal with respect to $(\cdot, \cdot)_2$. Hence

$$\varphi(l_1(x), ..., l_n(x)) = \psi(\xi_1(x), ..., \xi_n(x)),$$

where ψ is a Borel mapping on $\mathbb{R}^n$. Let us complement the collection $\xi_1, ..., \xi_n$ to an orthonormal basis $\{\xi_i\}$ in Λ. It is important that, with our convention about the versions of μ-measurable linear functionals, for every $t>0$, the functionals ξ_i are independent Gaussian random variables with respect to the Gaussian measure γ_t. Let $\{e_i\}$ be the standard orthogonal basis in l^2. Clearly, the vectors e_n are mutually orthogonal in X, $\|e_n\|_X = 1/n$. For every t, by the Ito–Nisio theorem (see [B98, Theorem 3.5.1; LT91]) one has

$$x = \sum_{i=1}^{\infty} \xi_i(x)\, e_i \tag{2.4}$$

for γ_t-almost all x, where the series converges in the norm of X. Indeed, the functionals ξ_i/t are independent standard Gaussian variables on (X, γ_t), and the vectors te_i form an orthonormal basis in the Cameron–Martin space of γ_t. Recall that this is l^2 with the inner product $t^{-2} \sum_{n=1}^{\infty} x_n y_n$. Thus, by (2.4) the transformation $x \mapsto (\xi_i(x))_{i=1}^{\infty}$ preserves the measure μ, and we have

$$\begin{aligned}\int_X \|x - \varphi(l_1(x), ..., l_n(x))\|^2\, \mu(dx) &= \int_X \|x - \psi(\xi_1(x), ..., \xi_n(x))\|^2\, \mu(dx)\\ &= \int_X \left\| \sum_{i=1}^{\infty} \xi_i(x)\, e_i - \psi(\xi_1(x), ..., \xi_n(x)) \right\|^2 \mu(dx)\\ &= \int_X \left\| \sum_{i=1}^{\infty} z_i e_i - \psi(z_1, ..., z_n) \right\|^2 \mu(dz),\end{aligned}$$

which will be shown to be infinite. Assume that this integral is finite. Let

$$\zeta(z_1, ..., z_n) = (\psi(z_1, ..., z_n), e_{n+1})_X, \qquad F(z) = z_{n+1} e_{n+1} - \zeta(z_1, ..., z_n)\, e_{n+1}.$$

Clearly, $F(z)$ is the orthogonal projection of $\sum_{i=1}^{\infty} z_i e_i - \psi(z_1, ..., z_n)$ to the subspace $\mathbb{R}^1 e_{n+1}$ in X. Therefore, $\int_X \|F(z)\|^2\, \mu(dz) < \infty$. Then, for almost each t, the corresponding integral with respect to γ_t is finite as well. Since the function $z \mapsto z_{n+1}$ is γ_t-square-integrable, we get the square-integrability of the function $z \mapsto \zeta(z_1, ..., z_n)$ with respect to γ_t for almost all t.

Note that for every t, by Fubini's theorem, which applies since γ_t is a product-measure, we have

$$\int_X z_{n+1} \zeta(z_1, ..., z_n)\, \gamma_t(dz) = 0.$$

Hence the nonnegative function $z_{n+1}^2 + \zeta(z_1, ..., z_n)^2$ is μ-integrable by the following calculation:

$$\begin{aligned}\int_X [z_{n+1}^2 + \zeta(z_1, ..., z_n)^2]\, \mu(dz) &= \int_0^\infty \int_X [z_{n+1}^2 + \zeta(z_1, ..., z_n)^2]\, \gamma_t(dz)\, p(t)\, dt \\ &= \int_0^\infty \int_X (z_{n+1} - \zeta(z_1, ..., z_n))^2\, \gamma_t(dz)\, p(t)\, dt \\ &= \int_X (n+1)^2\, \|F(z)\|^2\, \mu(dz) < \infty.\end{aligned}$$

This is a contradiction, since we have

$$\begin{aligned}\int_X z_{n+1}^2 \mu(dz) &= \int_0^\infty \int_X z_{n+1}^2 \gamma_t(dz)\, p(t)\, dt = \int_0^\infty \int_X t^2 z_{n+1}^2 \gamma(dz)\, p(t)\, dt \\ &= \int_X z_{n+1}^2 \gamma(dz) \int_0^\infty t^2 p(t)\, dt = \infty. \quad \blacksquare\end{aligned}$$

Note that the measure μ constructed in the previous example belongs to the class of the so-called l^2-orthogonally invariant measures, i.e., measures whose Fourier transforms are invariant under rotations on l^2. By Schönberg's theorem, in the infinite dimensional case, such measures without atoms at zero have the representations $\int_0^\infty \gamma^t \sigma(dt)$, where γ is the centered Gaussian measure with the Cameron–Martin space l^2 and σ is a probability measure on $(0, \infty)$ (see [VTC87, pp. 239, 245]). Measures of this type are discussed in [WW84].

EXAMPLE 2.3. If we choose p such that $\int_0^\infty t^r p(t)\, dt = \infty$, $\forall r > 0$, then the sane construction enables us to conclude that for any finite collection of μ-measurable linear functionals $l_1, ..., l_n$, and any Borel mapping φ: $\mathbb{R}^n \to X$, one has

$$\int_X \|x - \varphi(l_1(x), ..., l_n(x))\|^r\, \mu(dx) = \infty, \qquad \forall r > 0.$$

Proof. The proof is the same as above with the following technical modification. In order to derive the estimate $\int z_{n+1}^r \mu(dz) < \infty$ from $\int |z_{n+1} - \zeta(z_1, ..., z_n)|^r \, \mu(dz) < \infty$, we note that in fact

$$\int z_{n+1}^r \mu(dz) \leqslant \int |z_{n+1} - \zeta(z_1, ..., z_n)|^r \, \mu(dz),$$

due to the following estimate:

$$\int z_{n+1}^r \gamma_t(dz) \leqslant \int |z_{n+1} - \zeta(z_1, ..., z_n)|^r \, \gamma_t(dz), \qquad \forall t > 0. \tag{2.5}$$

To verify (2.5), it suffices to consider the projection ν of γ_t to $\mathbb{R}^{n+1}$. Then the desired estimate follows from the Fubini theorem applied to the product-measure ν and the following well-known and easily verified fact (see, e.g., [B98, Example 1.8.7; LT91]): if g is a centered Gaussian density on $\mathbb{R}^1$, then for every real ζ one has $\int_{-\infty}^{+\infty} |x - \zeta|^r \, g(x)\, dx \geqslant \int_{-\infty}^{+\infty} |x|^r \, g(x)\, dx$. ∎

The situation may be different in the case of *adaptive information* (see [WW84, TWW88]) when recursive inputs $N(x)$ are allowed; that is, $N(x)$ may have the form

$$\{l_1(x), l_2(l_1(x))(x), ..., l_{n(x)}(l_1(x), ..., l_{n(x)-1}(...))(x)\},$$

where for any k, $l_k(y_1, ..., y_{k-1})(x)$ is a function on $\mathbb{R}^{k-1} \times X$ which is a measurable linear functional in x and Borel in the other arguments (assuming, in addition, that substituting l_i for y_i we get a measurable function). Certainly, in this case it is reasonable to impose certain moment restrictions on the function n. However, even if recursive inputs are not admissible (i.e., the functions l_k do not depend on $(y_1, ..., y_{k-1})$), but the uniform boundedness of n is replaced by a weaker restriction like $\int n(x)^r \, \mu(dx) < \infty$, one can get a different picture. Let us recall the example constructed in [TW94, Theorem 6(1)]. Let σ_n be a nondecreasing sequence of positive numbers and let

$$D = \left\{ (x_n) : \sum_{n=1}^{\infty} \sigma_n^2 x_n^2 < \infty \right\}.$$

Define S: $D \to l^2$ by $Sx = \sum_{n=1}^{\infty} \sigma_n x_n e_n$, where $\{e_n\}$ is the standard basis of l^2. Finally, let μ be an atomic measure on $\{e_n\}$ such that $\sum_{n=1}^{\infty} \sigma_n^2 \mu(e_n) = \infty$. Clearly, μ has no finite second moment. As explained in [TW94], for any finite collection of linear functionals $l_i \in \mathbb{R}_0^\infty$ and any Borel map $\varphi\colon \mathbb{R}^n \to l^2$, one has

$$\int \|Sx - \varphi(l_1(x), ..., l_n(x))\|^2 \,\mu(dx) = \infty,$$

while if all μ-measurable linear functionals are admissible for l_i (i.e., all elements of $\mathbb{R}^\infty$), then one gets $Sx = \varphi(\lambda(x))$ μ-a.e., where $\lambda \in \mathbb{R}^\infty$ is represented by a sequence of distinct numbers η_n, $\varphi(\lambda(x)) = \sigma_n e_n$ if $\lambda(x) = \eta_n$ and $\varphi(\lambda(x)) = 0$ otherwise. In the terminology of [TW94] this means that the problem associated with S becomes solvable on the average for the class of functions $\Lambda = \mathbb{R}^\infty$ and it is unsolvable on the average for $\Lambda = \mathbb{R}_0^\infty$.

Now we observe that the problem becomes solvable even for $\Lambda = \mathbb{R}_0^\infty$ if inputs of varying cardinality with the moment restriction $\int n(x)^2 \,\mu(dx) < \infty$ are admissible. Indeed, one can put

$$l_n(x) = \sum_{i=1}^{M(n)} 2^{-i} x_i,$$

where $M(n)$ is a sequence of integers which increases sufficiently fast. Note that for any increasing sequence $\{n(e_k)\}$, one can find a sequence $\{M(n)\}$ such that $M(n(e_k)) \geqslant k$; this makes possible a choice of $n(e_k)$ such that $\int n(x)^2 \,\mu(dx) < \infty$. Then, the functionals $l_1, ..., l_{n(e_k)}$ uniquely determine the first k coordinates of $e_1, ..., e_k$. Therefore, the same φ as above can be chosen.

It would be interesting to investigate the Traub–Werschulz problem in the setting where the condition $\sup_{x \in X} n(x) < \infty$ is replaced by the weaker condition $\int n(x)^2 \,\mu(dx) < \infty$. It was shown in [WW84, Was86] that, in the case of a Gaussian measure or a mixture of homothetic images of a centered Gaussian measure, adaptive informations with uniformly bounded cardinality n are not more efficient than nonadaptive informations discussed above. The arguments used in [TWW88, WW84, Was86] to get this nice and surprising result yield the same conclusion in our example.

Example 2.4. In the situation of Example 2.2, one has

$$\int_X \|x - \varphi(l_1(x), l_2(l_1(x))(x), ..., l_n(l_1(x), ..., l_{n-1}(...)(x))(x))\|^2 \,\mu(dx) = \infty$$

for any algorithm φ and any adaptive information operator N (in the sense above) with any fixed cardinality n.

Proof. We shall explain how the same reasoning as in [TWW88, WW84, Was86], where continuous linear functionals and measures with finite second moments were considered, applies to our situation (where certain additional technical difficulties arise). The main idea of the proof is to show that the information operator N (possibly nonlinear) transforms the Gaussian measure γ to a Gaussian measure ν on $\mathbb{R}^n$ and, in addition, generates Gaussian conditional measures γ_y on the sets $N^{-1}(y)$. It is easy to reduce the general case to the case where, for every $y=(y_1, ..., y_n)\in\mathbb{R}^n$, the elements $l_1, l_2(y_1), ..., l_n(y_1, ..., y_n)$ are orthonormal in Λ_γ. To simplify notation, we shall write $l_j(y)$, $y=(y_1, ..., y_{j-1}, 0, ...)\in\mathbb{R}^n$, instead of $l_j(y_1, ..., y_{j-1})$. We have $N(x)=(g_1(x), ..., g_n(x))$, where

$$g_1(x)=l_1(x), \qquad g_2(x)=l_2(g_1(x))(x), \qquad g_j(x)=l_j(g_1(x), ..., g_{j-1}(x))(x).$$

It is possible to find a version of $x\mapsto l_j(y)(x)$ such that it is Borel and linear on a Borel linear subspace $D_{j,\,y}$ of full γ-measure measure. To this end, we observe that $l_j(y)(x)=\sum_{n=1}^{\infty}(x_n, l_j(y))_{L^2(\gamma)}\,x_n$ for γ-a.e. x, and the domain of convergence is Borel. In addition, for every y fixed, the domain of convergence in x is linear. Outside the domain of convergence of the series, we may put $l_j(y)(x)=0$. Let R_γ be the covariance operator of γ defined by the relation $f(R_\gamma g)=\int fg\,d\gamma$, $f\in X^*$, $g\in\Lambda_\gamma$ (due to a special structure of γ, if g is represented by an element of l^2, then $R_\gamma g$ is represented by the same element considered as a vector in X). Then $\gamma\circ N^{-1}$ is the standard Gaussian measure on $\mathbb{R}^n$. In addition, it is possible to choose Gaussian conditional measures γ_y, $y\in\mathbb{R}^n$, on X, i.e., every measure γ_y is concentrated on the set $N^{-1}(y)$, the function $y\mapsto\gamma_y(B)$ is measurable on $\mathbb{R}^n$ for every Borel set $B\subset X$, and

$$\gamma(B)=\int_{\mathbb{R}^n}\gamma_y(B)\,\gamma\circ N^{-1}(dy). \tag{2.6}$$

Finally, γ_y has mean $a^y=\sum_{j=1}^n y_j R_\gamma l_j(y)$ and covariance operator R^y: $f\mapsto R_\gamma f-\sum_{j=1}^n f(R_\gamma l_j(y))\,R_\gamma l_j(y)$. The first claim is verified by induction in n. Suppose it is true for $n-1$. Let us evaluate the Fourier transform of the measure $\gamma\circ N^{-1}$. This reduces to evaluating the integral $\int_X \exp(i[y_1g_1+\cdots+y_ng_n])\,d\gamma$. By the change of variables formula, it suffices to consider the case where $g_1(x)=x_1$. If we fix x_1 and integrate with respect to $(x_2, x_3, ...)$, then we get $\exp(iy_1x_1)\exp(-[y_2^2+\cdots+y_n^2]/2)$, which follows from the inductive assumption (recall that for x_1 fixed, $g_2=l_1(x_1)$ is a constant element of Λ_γ, $g_3(x)=l_3(x_1, g_2(x))(x)$, and so on). Integrating in x_1, we get $\exp(-[y_1^2+\cdots+y_n^2]/2)$, whence the first claim. Now let us prove the second claim. Note that by the above formula for $N(x)$, the sets $N^{-1}(y)$ are affine subspaces $N_y^{-1}(y)$ of codimension n, where N_y is the linear mapping to $\mathbb{R}^n$ generated by the functionals $l_1, l_2(y), ..., l_n(y)$ on the full measure Borel linear subspace $\bigcap_{j=1}^n D_{j,y}$. Clearly, $N_y^{-1}(y) = a^y + \operatorname{Ker} N_y$, since $N_y(a^y) = y$, which follows from the relationship $l_j(y)(R_\gamma l_k(y)) = (l_j(y), l_k(y))_{L^2(\gamma)} = \delta_{jk}$. In order to conclude that the announced measure γ_y is concentrated on $N^{-1}(y)$, it suffices to note that the integral of $\sum_{j=1}^n l_j(y)^2(x)$ with respect to the centered Gaussian measure with covariance operator R^y is zero. Finally, let us verify (2.6). To this end, denoting the Fourier transform of a measure λ by $\tilde{\lambda}$, it suffices to show that for every $f \in X^*$ one has

$$\tilde{\gamma}(f) = \int_{\mathbb{R}^n} \widetilde{\gamma_y}(f)\, \gamma \circ N^{-1}(dy).$$

By definition,

$$\widetilde{\gamma_y}(f) = \exp\left[i \sum_{j=1}^n y_j f(R_\gamma l_j(y)) \right] \exp\left[-\tfrac{1}{2} f(R_\gamma f) + \tfrac{1}{2} \sum_{j=1}^n f(R_\gamma l_j(y))^2 \right].$$

Let us integrate this expression in y with respect to the standard Gaussian measure on $\mathbb{R}^n$. Integrating first in y_n and using that the elements l_j do not depend on y_n, we get

$$\exp\left[i \sum_{j=1}^{n-1} y_j f(R_\gamma l_j(y)) - \tfrac{1}{2} f(R_\gamma f) + \tfrac{1}{2} \sum_{j=1}^n f(R_\gamma l_j(y))^2 \right] \exp[-\tfrac{1}{2} f(R_\gamma l_n(y))^2]$$
$$= \exp\left[i \sum_{j=1}^{n-1} y_j f(R_\gamma l_j(y)) \right] \exp\left[-\tfrac{1}{2} f(R_\gamma f) + \tfrac{1}{2} \sum_{j=1}^{n-1} f(R_\gamma l_j(y))^2 \right].$$

Integrating then in $y_{n-1}, ..., y_1$, we get $\exp(-f(R_\gamma f)/2) = \tilde{\gamma}(f)$, which proves our claim. In a similar way one proves that the image under N of the measure γ^t (the image of γ under $x \mapsto tx$) is the centered Gaussian measure on $\mathbb{R}^n$ with covariance matrix $t^2 I$ and that the conditional measures $(\gamma^t)_y$ are Gaussian with means a^y and covariance operators $t^2 R^y$. The rest of the proof is the same as in [TWW88, Appendix, Sect. 2.9.2; Was86, Theorem 4.2; WW84, Theorem 2.1] in a formally different situation (namely, for measures with finite second moments). Indeed, the integral in question can be written as

$$\int_0^\infty \int_{\mathbb{R}^n} \int_{N^{-1}(y)} \|x - \varphi(y)\|^2\, (\gamma^t)_y\,(dx)\, \gamma \circ N^{-1}(dy)\, p(t)\, dt.$$

Denoting the centered Gaussian measure with covariance operator $t^2 R^y$ by $\nu_{t,y}$, we rewrite the inner integral with respect to $(\gamma^t)_y$ as

$$\int_{\mathrm{Ker}\, N_y} \|x + a^y - \varphi(y)\|^2 \, \nu_{t,y}(dx) \geqslant \int_{\mathrm{Ker}\, N_y} \|x\|^2 \, \nu_{t,y}(dx) = t^2 c_n(y).$$

Integrating in y and t we get $+\infty$, whence the conclusion. ∎

3. MOMENTS OF MEASURES

Measures on infinite dimensional Banach spaces are typically concentrated on smaller linear subspaces. For example, the Wiener measure on $C[0, 1]$ is concentrated on Hölder subspaces H^α, $\alpha < 1/2$; more generally, any measure on $C[0, 1]$ is concentrated on the space of paths with some common modulus of continuity. This phenomenon is widely used in applications. Kuelbs [Ku73] found an abstract characterization of this phenomenon. He showed that for every measure μ on a separable Banach space X, one can always find a compactly embedded Banach space E of full measure (in fact, the construction of Kuelbs yields a dual Banach space). Ostrovskii [Ost80] used another method to construct a compactly embedded dual Banach space and investigated the moment conditions on this smaller space. Buldygin [Bul84] extended the result of Kuelbs and proved that E can be chosen separable reflexive. A further generalization is due to [B86], where a short proof has been given that the same is true for every Radon measure on a Fréchet space. However, for Radon measures on general locally convex spaces this assertion fails. Yet another natural and important possibility of extending Kuelb's result is to consider measures with finite moments or pre-Gaussian measures. Is it always possible to find E such that its norm is in $L^p(\mu)$ provided this holds true for the norm of X? Clearly, since $\|x\|_E \geqslant \mathrm{const}\, \|x\|_X$, not every E compactly embedded into X and having full measure satisfies this condition. For example, let μ on $X = l^2$ be defined by $\mu(e_n) = n^{-1}(\log(n+1))^{-2}$, where e_n is the nth standard basis vector in l^2. Then μ is pre-Gaussian and has all moments. Let $E = \{(x_n) \in l^2 : \|x\|_E^2 = \sum_{n=1}^\infty n^2 x_n^2 < \infty\}$. Then E is compactly embedded into X, $\mu(E) = \mu(X)$, $\int \|x\|_E^p \, \mu(dx) = \infty$ for all $p > 0$, and μ is not pre-Gaussian on E. The first positive result in this direction is due to Ostrovskii [Ost80] (see his Theorem 1 and Corollary 3), who proved that it is possible to find E in such a way that the moment restrictions are preserved. The following theorem extends the result cited and enables us to choose a separable reflexive E with the same property. Note that "reflexive" cannot be improved to "Hilbert." It is known [Sat76] that on every Banach space that is not linearly homeomorphic to a Hilbert space, there is a measure μ that vanishes on every continuously embedded Hilbert space. On many spaces, e.g., on $C[0,1]$, there are Gaussian measures μ with this property (see [B98, Sects. 3.6, 3.11]). Reflexive spaces are convenient in many respects, in particular, in connection with the approximation properties. Their main (and characteristic) advantage is the weak compactness of closed balls.

THEOREM 3.1. *Let X be a separable Banach space with a Borel probability measure μ possessing finite moment of order $r>0$. Then there is a linear subspace E of X with the following properties*:

(i) *E with some norm $\|\cdot\|_E$ is a separable reflexive Banach space, the unit ball of which is compact in X*;

(ii) *$\mu(E)=1$ and $\int_E \|z\|_E^r\,\mu(dz)<\infty$.*

If μ has finite second moment and is pre-Gaussian, then E can be chosen in such a way that μ has finite second moment and remains pre-Gaussian on E. Finally, if μ has all moments on X, then E can be chosen with the same property.

Proof. We need the following technical result. Let φ be a decreasing nonnegative function on $[0,\infty)$ such that $\sum_{n=1}^\infty \varphi(n)<\infty$. Then there is a sequence of positive numbers α_n decreasing to zero such that $\alpha_n n\to\infty$ and $\sum_{n=1}^\infty \varphi(\alpha_n n)<\infty$. Indeed, there exists a sequence of natural numbers C_n increasing to infinity such that the series $\sum_{n=1}^\infty C_n\varphi(n)$ converges. Put

$$S_n=\sum_{j=1}^n C_j,\qquad S_0=0,\qquad \beta_n=S_n/(n+1).$$

Let $\alpha_j=\beta_n^{-1}$ if $S_n\leqslant j<S_{n+1}$. Then $\beta_n\to\infty$, since $C_n\to\infty$. Hence $\alpha_n\to 0$. Clearly, $\alpha_j\,j\geqslant n+1$ if $S_n\leqslant j<S_{n+1}$. Hence $\alpha_n n\to\infty$. In addition,

$$\sum_{S_n\leqslant j<S_{n+1}} \varphi(a_j\,j)\leqslant \sum_{S_n\leqslant j<S_{n+1}} \varphi(\alpha_j S_n)=\sum_{S_n\leqslant j<S_{n+1}} \varphi(n+1)\leqslant C_{n+1}\varphi(n+1).$$

This estimate implies the convergence of the series $\sum_{n=1}^\infty \varphi(\alpha_n n)$.

Let us return to the proof of our main claim. Let $\varphi(R)=\mu(x:\|x\|>R^{1/r})$. By the integrability of $\|\cdot\|^r$, we get the convergence of the series $\sum_{n=1}^\infty \mu(x:\|x\|>n^{1/r})$. As we have proved above, there is a sequence of positive numbers α_n decreasing to zero with $\alpha_n n\to\infty$ such that the series $\sum_{n=1}^\infty \varphi(\alpha_n n)$ converges. For every n, let us choose a compact set K_n in the centered ball U_n of radius $n^{1/r}$ such that

$$\mu(\alpha_n^{1/r}K_n)\geqslant \mu(\alpha_n^{1/r}U_n)-2^{-n}. \tag{3.7}$$

The set $K=\bigcup_{n=1}^{\infty} c_n K_n$, $c_n:=\alpha_n^{1/r} n^{-1/r}$, is relatively compact. Indeed, for any sequence $\{x_n\}$ in this set, either its infinite subsequence is contained in one of the $c_n K_n$'s or the whole sequence converges to zero. It is well known that the closed balanced convex hull V of K is also compact (see [S71, Corollary 4.3, Sect. 5, Chap. II, p. 50]). Let p_V be the corresponding gauge functional defined on the linear span of V by

$$p_V(x)=\inf\{\lambda>0 : x/\lambda\in V\}.$$

We shall verify the inclusion $p_V\in L^r(\mu)$ (note that the functional p_V is measurable since $\{p_V\leqslant c\}=cV$ for $c\geqslant 0$). The set $\{x : p_V(x)\leqslant n^{1/r}\}=n^{1/r}V$ contains $n^{1/r}K$; hence it contains also $\alpha_n^{1/r}K_n$. Therefore, by virtue of (3.7) we arrive at the estimate

$$\begin{aligned}\mu(x : p_V^r(x)>n)&=1-\mu(x : p_V(x)\leqslant n^{1/r})\leqslant 1-\mu(\alpha_n^{1/r}K_n)\\&\leqslant 1+2^{-n}-\mu(\alpha_n^{1/r}U_n)=2^{-n}+\mu(x : \|x\|>\alpha_n^{1/r}n^{1/r})\\&=2^{-n}+\varphi(\alpha_n n),\end{aligned}$$

whence $p_V^r\in L^1(\mu)$. By construction, the linear span of V has full measure. Indeed, it contains the sets $\alpha_n^{1/r}K_n$. By (3.7), $\mu(\alpha_n^{1/r}K_n)\to 1$, since due to the condition $\alpha_n n\to\infty$, the $\alpha_n^{1/r}U_n$'s are the balls of radii $(\alpha_n n)^{1/r}\to\infty$. Note that instead of the argument above we could use the construction sketched in [Ost80]. However, the reasoning given above applies to seminorms on Fréchet spaces; on the other hand, it can be extended to a more general situation considered in [Ost80, Theorem 1] with certain increasing functions of the norm replacing powers of the norm.

The next step is to prove the existence of a separable reflexive Banach space E compactly embedded into X and containing V as a bounded set. This will imply that $\mu(E)=1$ and that on the linear span of V the norm $\|\cdot\|_E$ is majorized by const p_V, whence property (ii). It should be noted that by a classical result, all Borel subsets of E are Borel in X (see, e.g., [Sch73]); hence μ can be restricted to the Borel σ-field of E. It is important that the linear span E_V of V endowed with the norm p_V is a Banach space with unit ball V (see [S71, p. 97]) and hence it is compactly embedded into X. According to [S71, Lemma 1, Chap. III, Sect. 9, p. 111] there is an absolutely convex compact subset A in X containing V such that V is compact as a subset of the Banach space E_A with the norm p_A generated by A as explained above. Let us apply the same result to A and find a yet bigger absolutely convex compact C such that A is compact as a subset of the associated Banach space E_C. Now, by the factorization lemma of W. J. Davis, T. Figiel, W. B. Johnson, and A. Pelczynski (see [D75, Lemma in Sect. 4, Chap. IV, p. 160]), there exists a reflexive Banach space Z with

$E_A \subset Z \subset E_C$ such that Z is continuously embedded into E_C and, in addition, the unit ball of E_A is bounded in Z, i.e., E_A is continuously embedded into Z. Therefore, the set V is compact in the Banach space Z. Hence the closure of the linear span of V in the norm of Z is a separable reflexive Banach space (with the norm from Z) which we take for E. Clearly, V is bounded in E. In addition, the embedding $E \to E_C$ is continuous; hence by the compactness of the embedding $E_C \to X$, the space E is compactly embedded into X. Note that the closed unit ball of E is compact in X (and not just relatively compact). Indeed, since E is reflexive, its closed balls are weakly compact; hence they are weakly compact also in X due to the continuity of the embedding $E \to X$ for the weak topologies (see [S71, Sect. 2, Chap. IV, p. 129]). Being convex, they are then closed in the norm topology of X (see [S71, Sect. 3, Chap. IV, p. 130]) and hence compact in X. Note that the idea to apply the factorization lemma to get a reflexive support is due to [Bul84].

If μ has all strong moments, then it suffices to take $\alpha_n = \log(n+1)$ in the construction above applied to the case $p=1$. Finally, if μ is pre-Gaussian, then in order to preserve this property on E we apply the construction above to the measure $v = (\mu + \gamma)/2$, where γ is the corresponding Gaussian measure on X. Therefore, we get a separable reflexive Banach space E compactly embedded into X and having full measure with respect to v such that v has finite second moment on E. We have to check that for all $f, g \in E^*$ one has

$$\int f(x)\,\mu(dx) = \int f(x)\,\gamma(dx), \qquad \int f(x)\,g(x)\,\mu(dx) = \int f(x)\,g(x)\,\gamma(dx).$$

By condition, this is true for all elements in X^*. Now let $f, g \in E^*$. Applying the same result to v on E, we get a separable Banach space E_0 of full v-measure whose closed unit ball B is compact in E and, in addition, v has finite second moment on E_0. By virtue of [S71, Lemma 2, Chap. III, Sect. 9, p. 112], there exist two sequences $\{f_n\}$ and $\{g_n\}$ in X^* convergent uniformly on B to f and g, respectively. Since $|f_n(x) - f(x)| \leqslant \sup_B |f - f_n|\, p_B(x)$ for all $x \in E_0$ (that is, v-a.e.) and $p_B \in L^2(v)$, we conclude that $f_n \to f$ both in $L^2(\mu)$ and $L^2(\gamma)$. The same is true for g_n and g. Hence, we arrive at the desired identity. ∎

Recall that a Banach space X is said to have the approximation property if for every compact set $K \subset X$ and every $\varepsilon > 0$, there is a continuous linear operator T with finite dimensional range such that $\|x - Tx\| \leqslant \varepsilon$ for all $x \in K$. It is known that not every separable Banach space has this property (see, e.g., [Sin81]). Our next result links this section with the previous one (cf. Proposition 2.1).

COROLLARY 3.2. *Let μ be a probability measure on X with finite second moment. Assume that X has the approximation property. Then for every $\varepsilon > 0$ there is a continuous linear operator T with finite dimensional range such that*

$$\int_X \|x - Tx\|^2 \, \mu(dx) < \varepsilon.$$

Proof. Let E be as in the previous theorem and let K be its unit ball. Let $\varepsilon_0 > 0$ be such that $\varepsilon_0 \int_E \|z\|_E^2 \, \mu(dz) < \varepsilon$. Take a finite dimensional operator T with $\sup_K \|z - Tz\| \leqslant \varepsilon_0$. Then $\|z - Tz\| \leqslant \varepsilon_0 \, \|z\|_E$ for all $z \in E$. Hence $\int_E \|z - Tz\|^2 \, \mu(dz) \leqslant \varepsilon_0 \int_E \|z\|_E^2 \, \mu(dz) < \varepsilon$. ▮

This corollary generalizes Theorem 17 from [M90], where the space X was assumed to have the bounded approximation property (i.e., there exists $\lambda > 0$ such that for every $\varepsilon > 0$ and every compact set $K \subset X$ there exists a finite rank operator L on X with $\|L\| \leqslant \lambda$ and $\sup \{ \|x - Lx\|, x \in K\} \leqslant \varepsilon$), which is strictly stronger than the approximation property. We do not know whether one can always take for E in Theorem 3.1 a space with the approximation property. It is also open whether the assertion of the last corollary holds true for arbitrary Banach spaces.

Remark 3.3. Let μ be a probability measure on a separable Banach space X having finite second moment. There is a sequence $\{l_n\} \subset X^*$ generating the Borel σ-field of X (see [VTC87, Chap. I]). Denote by $\mathscr{A}_n$ the σ-field generated by $l_1, \ldots, l_n$. It follows from the vector martingale convergence theorem (see, e.g., [VTC87, Theorem 4.1 in Sect. II.4.2, p. 128, and Theorem 4.2 in Sect. II.4.3, p. 131]) that the sequence $\{\xi_n\}$ of the conditional expectations with respect to σ_n of the identity map I converges to I in $L^2(\mu, X)$ and μ-a.e. Thus, we get Borel mappings $\varphi_n \colon \mathbb{R}^n \to X$ such that $\int \|x - \varphi_n(l_1(x), \ldots, l_n(x))\|^2 \, \mu(dx) \to 0$. However, these mappings need not be linear as in Corollary 3.2 or continuous as in Proposition 2.1.

Remark 3.4. As shown by Carmona [C77], for any Gaussian vector Y with distribution μ in a separable Banach space X, any orthonormal basis $\{e_i\}$ in the reproducing kernel Hilbert space H of μ, and any sequence of μ-measurable linear functionals g_i that are independent standard Gaussian random variables with respect to μ, the sequence $P_n(\omega) = \sum_{i=1}^n g_i(Y(\omega)) \, e_i$ converges to Y in any L^p-space of X-valued maps. This means that if $Y(x) = Ax$, where A is a μ-measurable linear operator, then one gets L^p-approximations of Ax by finite-dimensional linear operators. This fact was noted later also in other papers (see, e.g., references in [TW94]). Clearly, this fact is an immediate corollary of Fernique's theorem on the exponential integrability and the martingale convergence theorem (with an even stronger conclusion about the convergence of the exponential moments). It would be interesting to investigate this problem for wider classes of measures, e.g., for log-concave measures (which in many aspects are similar to Gaussian measures). Recall that a probability measure μ is called log-concave if

$$\mu(\lambda A + (1-\lambda) B) \geqslant \mu(A)^{\lambda} \mu(B)^{1-\lambda}, \qquad \forall A, B \in \mathscr{B}(X), \qquad \lambda \in [0, 1].$$

Gaussian measures are known to be log-concave. If $X = \mathbb{R}^n$, then μ is log-concave if and only if it is concentrated on some affine subspace of $\mathbb{R}^n$ and admits density $\exp(-V)$ with respect to the corresponding Lebesgue measure with a convex function V. It is shown by Borell [Bor74] that if μ is log-concave, then for every μ-measurable seminorm q there is $c > 0$ such that $\exp(cq) \in L^1(\mu)$. In particular, if F is a μ-measurable linear operator, then $\|Fx\| \in L^2(\mu)$. Hence, for spaces X with a Schauder basis or with the approximation property, the results above apply and yield the existence of finite dimensional approximations of F in $L^2(\mu, X)$. It is not clear whether such linear approximations exist without additional assumptions about X.

REFERENCES

[B86] V. I. Bogachev, Locally convex spaces with the CLT property and supports of measures, *Moscow Univ. Math. Bull.* **41** (1986), 19–23.

[B98] V. I. Bogachev, "Gaussian Measures," Amer. Math. Soc., Providence, RI, 1998.

[Bor74] C. Borell, Convex measures on locally convex spaces, *Ark. Math.* **12** (1974), 239–252.

[Bor76] C. Borell, Gaussian Radon measures on locally convex spaces, *Math. Scand.* **38** (1976), 265–284.

[Bul84] V. V. Buldygin, Supports of probability measures in separable Banach spaces, *Theory Probab. Appl.* **29** (1984), 546–549.

[C77] R. Carmona, Measurable norms and some Banach space valued Gaussian processes, *Duke Math. J.* **44** (1977), 109–127.

[D75] J. Diestel, Geometry of Banach spaces, *in* "Lecture Notes in Mathematics," Vol. 485, Springer-Verlag, Berlin/New York, 1975.

[H90] S. Heinrich, Probabilistic complexity analysis for linear problems in bounded domains, *J. Complexity* **6** (1990), 231–255.

[K78] M. Kanter, Random linear functionals and why we study them, *in* "Lecture Notes in Mathematics," Vol. 645, pp. 114–123, Springer-Verlag, Berlin/New York, 1978.

[Ku73] J. Kuelbs, Some results for probability measures on linear topological vector spaces with an application to Strassen's LogLog Law, *J. Functional Anal.* **14** (1973), 28–43.

[LT91] M. Ledoux and M. Talagrand, "Probability in Banach Spaces. Isoperimetry and Processes," Springer-Verlag, Berlin/New York, 1991.

[M90] P. Mathé, s-Numbers in information-based complexity, *J. Complexity* **6** (1990), 41–66.

[Ost80] E. I. Ostrovskii, On the support of probability measures in separable Banach spaces, *Dokl. Akad. Nauk SSSR* **225** (1980), 1319–1320 [In Russian; English transl.: *Soviet Math. Dokl.* **22** (1980), 836–838].

[Sat76] H. Sato, Banach support of a probability measure in a locally convex space, *in* "Lecture Notes in Mathematics," Vol. 526, pp. 221–226, Springer-Verlag, Berlin/New York, 1976.

[S71] H. H. Schaefer, "Topological Vector Spaces," Springer-Verlag, Berlin/New York, 1971.

[Sch73] L. Schwartz, "Radon Measures on Arbitrary Topological Spaces and Cylindrical Measures," Oxford Univ. Press, London, 1973.

[Sin81] I. Singer, "Bases in Banach spaces. II," Springer-Verlag, Berlin/New York, 1981.

[Sm83] W. Smolenski, An abstract form of a counterexample of Marek Kanter, *in* "Probability Theory on Vector Spaces, III (Lublin, 1983)," Lecture Notes in Mathematics, Vol. 1080, pp. 288–291, Springer-Verlag, Berlin/New York, 1983.

[TWW88] J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, "Information-based Complexity," Academic Press, New York, 1988.

[TW94] J. F. Traub and A. G. Werschulz, Linear ill-posed problems are solvable on the average for all Gaussian measures, *Math. Intelligencer* **16** (1994), 42–48.

[VTC87] N. N. Vakhania, V. I. Tarieladze, and S. A. Chobanyan, "Probability Distributions in Banach Spaces," Kluwer Academic, Dordrecht/Norwell, MA, 1987.

[V94] L. Veselý, Renorming a Banach space to make a given subspace Chebyshev, *Arch. Math.* **62** (1994), 254–260.

[Was86] G. Wasilkowski, Optimal algorithms for linear problems with Gaussian measures, *Rocky Mountain J. Math.* **16** (1986), 727–749.

[WW84] G. Wasilkowski and H. Woźniakowski, Can adaption help on the average? *Numer. Math.* **44** (1984), 169–190.

[Wer87] A. G. Werschulz, What is the complexity of ill-posed problems? *Numer. Functional Anal. Optim.* **9** (1987), 945–967.

[Wer91] A. G. Werschulz, "The Computational Complexity of Differential and Integral Equations: An Information-based Approach," Oxford Univ. Press, Oxford, 1991.